

Title	Go-and-Back method: effective estimation of the hidden motion of proteins from single-molecule time series.
Author(s)	Miyazaki, Makito; Harada, Takahiro
Citation	The Journal of chemical physics (2011), 134(13)
Issue Date	2011-04-07
URL	http://hdl.handle.net/2433/141839
Right	© 2011 American Institute of Physics
Type	Journal Article
Textversion	publisher

Go-and-Back method: Effective estimation of the hidden motion of proteins from single-molecule time series

Makito Miyazaki^{1,a)} and Takahiro Harada^{2,b)}

¹*Department of Physics, Graduate School of Science, Kyoto University, Kyoto 606-8502, Japan*

²*Department of Physics, Graduate School of Science, The University of Tokyo, Tokyo 113-0033, Japan*

(Received 17 November 2010; accepted 16 March 2011; published online 7 April 2011)

We present an effective method for estimating the motion of proteins from the motion of attached probe particles in single-molecule experiments. The framework naturally incorporates Langevin dynamics to compute the most probable trajectory of the protein. By using a perturbation expansion technique, we achieve computational costs more than 3 orders of magnitude smaller than the conventional gradient descent method without loss of simplicity in the computation algorithm. We present illustrative applications of the method using simple models of single-molecule experiments and confirm that the proposed method yields reasonable and stable estimates of the hidden motion in a highly efficient manner. © 2011 American Institute of Physics. [doi:10.1063/1.3574396]

I. INTRODUCTION

In single-molecule experiments of molecular motors, it has been a widely adopted strategy to visualize continuous stepwise motion by attaching a large probe particle.^{1–5} Recently, this technique has also been put into use for monitoring conformational changes in proteins that stochastically switch between two or more metastable states.^{6,7} Compared to using a fluorescent dye, using a probe particle has the following advantages: First, advances in technology now enable the monitoring of the particle at ultrahigh temporal and spatial resolutions of up to 9.1 μ s (Ref. 8) and 0.1 nm,^{5,9} respectively. Second, the particle can be manipulated under optical microscopes, which provides insights into single-molecule mechanics^{3,10–13} and energetics.^{14,15} However, one problem that remains in this method is that the observed motion of the probe particle does not precisely reflect the protein motion. For typical experiments, since the probe is large and loosely connected to the protein, the motion of the probe is usually delayed. To study protein dynamics in detail, the motion and the physical parameters of the protein from the observed trajectory of the probe particle must be estimated.

To date, there has been systematic effort in developing both theoretical and numerical frameworks to determine a discrete state model of proteins from single-molecule fluorescent spectroscopy.^{16–21} The framework was recently combined with Bayesian statistics,²² which allows to analyze the entire sequence of single-molecule data, and it was experimentally demonstrated that the method yields more reliable estimation than the conventional correlation analysis.²³ By the use of entire time series, a method to extract effective energy landscape has also been developed recently.^{24,25}

For the time-series analysis of probe particles, there are several numerical approaches to estimating the underlying

stepwise trajectories of the molecular motors from the motion of probe particles, and these have been applied to various kinds of experiments.^{26–30} However, to the best of authors' knowledge, all of these approaches attempt to discretize the observed trajectories, i.e., the probe trajectories into several discrete states. More importantly, these approaches do not incorporate the dynamics of the entire system, namely, thermal fluctuations of both the protein and the probe, and the response delay of the probe motion. Instead, a sampling technique of the reaction pathway in continuous space, so-called transition path sampling,^{31,32} is based on Langevin dynamics. However, this requires a significant computational cost to search for the dominant pathways, making it inefficient in the presence of multiple reaction pathways.³³ Although an effective method of estimating dominant reaction pathways (DRPs) has recently been developed,^{33–35} it is not straightforwardly applicable to the analysis of time series data because the method performs the path sampling using not constant time steps but constant displacement steps.

In the present article, we consider a Langevin system that consists of two Brownian particles (one is visible and the other is hidden) connected with each other. On the basis of this model, we propose a method to efficiently estimate DRPs of the hidden variable from the trajectories of the visible variable. Although the model is very simple, it can be considered as a crude description of single-molecule experimental setups under appropriate approximations. We assume that the model and all the parameters have been determined and focus on the estimation of the DRP of the hidden variable. As will be shown later, even though the model is already given, finding the most probable trajectory of the hidden variable using the trajectory of the visible variable remains nontrivial. For parameter estimation in the presence of hidden degrees of freedom, we have proposed a general framework and the practical utility was demonstrated by a simple Langevin model.³⁶ In this framework, the DRP plays a central role in parameter estimation. Our final remarks in the present article are a discussion of practical application of the proposed method to parameter estimation.

^{a)} Also at Department of Physics, Graduate School of Science, the University of Tokyo, Tokyo, Japan. Electronic mail: miyazaki@chem.scphys.kyoto-u.ac.jp.

^{b)} Electronic mail: harada@phys.s.u-tokyo.ac.jp.

Once the model is given, the path-probability can be expressed in terms of the Onsager–Machlup path probability.^{37–39} Thus, in principle, we can apply a standard maximum likelihood estimation to the hidden trajectory. However, we find in general that the conventional optimization algorithm requires too high a computational cost to find the DRP. Here, we develop an effective approximation technique with the aid of perturbation theory. The schematic procedure of our method is as follows. First, on the basis of a rough estimate of the DRP, we solve one differential equation in the forward-time direction. Next, by substituting this solution, we solve another differential equation in the reverse-time direction and obtain a better estimate of the DRP. By repeating this procedure, we can systematically increase the accuracy of the estimate. Since we solve these differential equations by alternating between the forward and reverse (backward) directions, we name this algorithm the Go-and-Back method.

In Sec. II, we introduce a working model and discuss the validity of the model. Then, we explain the problem with the gradient descent method and derive the Go-and-Back method. In Sec. III, we examine two models of typical single-molecule experiments to investigate the effectiveness of our method.

II. FRAMEWORK

A. Model

The molecular structure of protein typically consists of a few hundred of amino acids. Therefore, in general, we need to consider such a huge number of degrees of freedom to study molecular dynamics of proteins. However, recent experimental^{6,7,15,40,41} studies on motor proteins clarified that only a few degrees of freedom dominate the large-scale conformational changes. In a particular case of the rotational molecular motor F₁-ATPase, it has been experimentally shown that the energy conservation of the entire system is explained by considering only one-dimensional (rotational) motion.^{15,41} In addition, a normal mode analysis on motor proteins^{42–45} implies that the low-frequency modes (a few degrees of freedom) correspond to such large-scale motion (μ s–ms), and the low-frequency modes are distinct from higher frequency modes (a huge number of degrees of freedom) that may correspond to the local conformational fluctuations and the catalytic reactions at the active site (ps–ns). Therefore, within the typical time resolution of an optical microscope (μ s–ms), the large number of high-frequency modes are eliminated from the dynamics of proteins and thus the dynamics can be approximated by low-dimensional overdamped Langevin equations.

Taking into account the above facts, we consider a Langevin system that consists of two Brownian particles interacting with each other:

$$\gamma \dot{x} = -\partial_x [V^{\text{eff}}(x) + U(x, y)] + \xi, \quad (1)$$

$$\Gamma \dot{y} = F(y) - \partial_y U(x, y) + \eta, \quad (2)$$

where γ and Γ are the friction coefficients, $\xi(t)$ and $\eta(t)$ are zero-mean white Gaussian noise with variances $2\gamma k_B T$ and $2\Gamma k_B T$, respectively. If $x(t)$ and $y(t)$ are regarded as the dominant degrees of freedom of the protein and the probe particle, respectively, then Eqs. (1) and (2) can be considered as a crude model of single-molecule experiments.^{36,46,47} Note that we consider the simplest case where both $x(t)$ and $y(t)$ are one-dimensional variables because typical single-molecule experiments monitor only one-dimensional motion, but the following calculation is straightforwardly extended to higher dimensional $x(t)$ and $y(t)$.

In the present model, γ corresponds to the sum of the internal friction coefficient of the protein and the viscous friction coefficient between the protein and the medium. Γ corresponds to the viscous friction coefficient of the probe particle. For simplicity, we assume that γ and Γ are position-independent. $U(x, y)$ corresponds to the energy potential of the elastic linker between the protein and the probe particle. $V^{\text{eff}}(x) \equiv V(x) - fx$ is the effective energy potential of the protein, where $V(x)$ corresponds to the energy landscape profile of the protein along the reaction coordinate and f corresponds to the “driving force” provided by the catalytic reaction such as ATP hydrolysis. In actual experiments, trapping force or space-constant load is sometimes applied to the probe particle. We also incorporate such an external force into the model equations denoted by $F(y)$. We assume that $U(x, y)$, $V^{\text{eff}}(x)$, and $F(y)$ are independent of t .

Here, let us discuss the validity of the working model by using two examples displayed in Fig. 1. First, we consider that a protein has two chemical states: A and B states, and the protein stochastically goes back and forth between the two states. We suppose that the two chemical states exhibit different conformations and we can observe the switching motion by

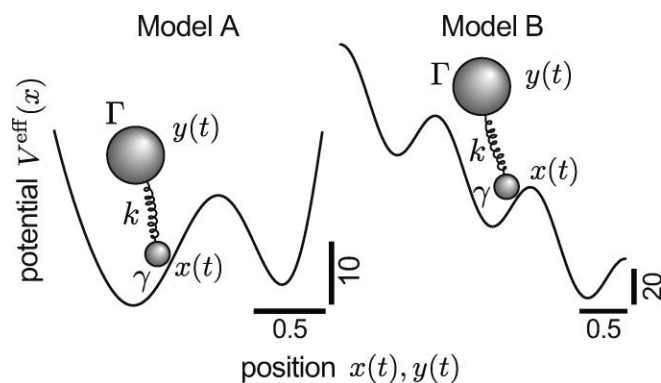


FIG. 1. Models for the numerical experiments. Model A: $x(t)$ is stochastically switching between two local minima of $V^{\text{eff}}(x)$ by means of thermal noise (equilibrium state). Model B: $x(t)$ is stochastically stepping down the tilted effective potential $V^{\text{eff}}(x) \equiv V(x) - fx$, where $V(x)$ is a periodic function and f is the driving force (nonequilibrium steady state). The parameters in these models are chosen as follows. Model A: $V^{\text{eff}}(x)$ is a polynomial function defined as $V^{\text{eff}}(x) \equiv \sum_{i=1}^8 a_i x^i$, where $a_1 = 2.57$, $a_2 = 114.4$, $a_3 = -14.0$, $a_4 = -191.9$, $a_5 = -81.3$, $a_6 = 196.4$, $a_7 = 29.1$, and $a_8 = -51.5$. (In typical experiments, the bimodal distribution of the probe particle is best fitted by two Gaussian functions. Thus, we roughly fitted two Gaussian functions of different means and variances with an eighth-order polynomial function in order to reproduce experimental results.) Model B: $V(x) \equiv A \sin(2\pi x/l)$, where $A = 20$, $l = 1$, and $f = 40$. In both models, $\gamma = 2$, $\Gamma = 20$, $k = 160$, $k_B T = 4.11$, and $\Delta t = 0.01$.

attaching a probe particle. Here, if the off rates of A(substrate) and B(product) from the protein are slower than the switching rates between the two states and the observation period, the entire system is well approximated as an equilibrium state. In addition, if the switching rates are slower than the global relaxation timescale of the protein structure, namely, the transition rates are well explained by the Kramers model,⁴⁸ the entire dynamics can be modeled by a double-well potential of $V(x)$ with $f = 0$ and $F = 0$ [Fig. 1(a)].

Next, we suppose that a molecular motor translocates in one direction with regular steps by catalytic reactions. For example, a rotary molecular motor F_1 -ATPase rotates counter-clockwise with regular 120° steps by hydrolyzing ATP.² In typical observation period (several minutes), the entire system is regarded as a nonequilibrium steady state because the concentration of ATP and the products (ammonium dihydrogen phosphate and P_i) are almost constant in this period. If ATP molecules are abundant in the medium and the rate limiting reaction of each step is the global conformational change of the protein by its thermal fluctuation, the phenomenological model can be described by using a tilted periodic potential [Fig. 1(b)].

In this manner, the dynamics of proteins and attached probe particles can be approximated as the simple Langevin model under appropriate conditions. Of course, the present model has several limitations to apply actual experiments due to the simple approximations especially the position-independent γ and the time-independent $V^{\text{eff}}(x)$. The details will be discussed in Sec. IV.

In what follows, we assume that $y(t)$ is observed with a sufficiently high temporal and spacial resolution, while $x(t)$ is hidden. We also assume that the entire set of system parameters is given. Then, our final task to consider here is the estimation of the most probable trajectory of $x(t)$ from the trajectory of $y(t)$.

B. Path probability

We denote the set of the trajectories from time $t = 0$ to $t = \tau$ as $[x, y]$ and the entire set of system parameters as $\mathbf{\Pi} = (\Pi_1, \Pi_2, \dots, \Pi_p)$. Given a value of $\mathbf{\Pi}$, we can calculate the path probability $P([x, y]|\mathbf{\Pi})$ as follows.

First, $P([x, y]|\mathbf{\Pi})$ is decomposed into the initial and transition probabilities as

$$P([x, y]|\mathbf{\Pi}) = P_{\text{init}}(x_0, y_0|\mathbf{\Pi})P_{\text{tr}}((x_0, y_0) \rightarrow [x, y]|\mathbf{\Pi}), \quad (3)$$

where x_0 and y_0 denote $x(0)$ and $y(0)$, respectively.

Next, the transition probability can be expressed in terms of the Onsager–Machlup path-probability:^{37–39}

$$P_{\text{tr}}((x_0, y_0) \rightarrow [x, y]|\mathbf{\Pi}) = C \exp[-\beta S([x, y]|\mathbf{\Pi})], \quad (4)$$

where $\beta^{-1} \equiv k_B T$ and the action functional $S([x, y]|\mathbf{\Pi})$ is

defined as

$$\begin{aligned} S([x, y]|\mathbf{\Pi}) \equiv & \frac{1}{4\gamma} \int_0^\tau [\gamma \dot{x} + V_x^{\text{eff}}(x) + U_x(x, y)]^2 dt \\ & - \frac{k_B T}{2\gamma} \int_0^\tau [V_{xx}^{\text{eff}}(x) + U_{xx}(x, y)] dt \\ & + \frac{1}{4\Gamma} \int_0^\tau [\Gamma \dot{y} - F(y) + U_y(x, y)]^2 dt \\ & - \frac{k_B T}{2\Gamma} \int_0^\tau [-F_y(y) + U_{yy}(x, y)] dt, \quad (5) \end{aligned}$$

where the total and partial differentiations are denoted using the same notation such as $V^{\text{eff}} \equiv V_x^{\text{eff}}$ and $\partial_{xx} U \equiv U_{xx}$. When the trajectory $[x, y]$ is time-discretized by Δt , the normalization constant becomes $C = [\sqrt{\gamma\Gamma}/(4\pi k_B T \Delta t)]^N$, where $\tau \equiv N \Delta t$.

Therefore, if we adopt an appropriate approximation for the initial distribution $P_{\text{init}}(x_0, y_0|\mathbf{\Pi})$ (Ref. 49), we can compute the path probability by Eqs. (3)–(5).

Once we obtain a concrete expression for the path probability, we can estimate the DRP by a standard maximum likelihood estimation with respect to $[x]$. In Bayesian statistics, the maximum likelihood estimator (MLE) is included in the maximum *a posteriori* (MAP) estimator as a special case. It has been clarified in our previous work that the MAP estimator does not coincide with the true trajectory of $x(t)$.³⁶ However, when the motion of $x(t)$ is stepwise similar to that of molecular motors, we find that the MAP estimator seems to be a good estimator of the stepping motion of $x(t)$. To maintain consistency with our previous work and also for convenience when considering the application of the Go-and-Back method to parameter estimation, we refer to the MLE as the MAP estimator in the present article, which we denote by $[\hat{x}]$ in the following sections.

C. Gradient descent

The gradient descent method is the most widely used optimization algorithm. Before proceeding to the Go-and-Back method, let us consider why this standard method is inefficient for the present problem.

To maximize $P([x, y]|\mathbf{\Pi})$ with respect to $[x]$, an initial condition for $[x]$ and a boundary condition are required. Here, we adopt the Dirichlet boundary condition. (For the initial condition, for instance we can adopt $[x] = [y]$.) In this case, x_0 is fixed, and thus we avoid having to consider the initial distribution $P_{\text{init}}(x_0, y_0|\mathbf{\Pi})$. Therefore, the maximization of $P([x, y]|\mathbf{\Pi})$ is replaced by the minimization of action functional $S([x, y]|\mathbf{\Pi})$ with respect to $[x]$. By introducing a “virtual time” s and replacing $x(t)$ with $x(t, s)$, we obtain the MAP estimator $[\hat{x}]$ as the solution of the following partial differential equation in the limit of $s \rightarrow \infty$:

$$\begin{aligned} \frac{\partial x(t, s)}{\partial s} &= - \frac{\delta S([x, y]|\mathbf{\Pi})}{\delta x} \\ &= - \left. \frac{\partial W(x, y)}{\partial x} \right|_{x=x(t, s)} + \frac{\gamma}{2} \frac{\partial^2 x(t, s)}{\partial t^2}, \quad (6) \end{aligned}$$

where

$$W(x, y) \equiv \frac{1}{4\gamma} [V_x^{\text{eff}}(x) + U_x(x, y)]^2 - \frac{k_B T}{2\gamma} [V_{xx}^{\text{eff}}(x) + U_{xx}(x, y)] + \frac{1}{4\Gamma} [-F(y) + U_y(x, y)]^2 - \frac{k_B T}{2\Gamma} [-F_y(y) + U_{yy}(x, y)]. \quad (7)$$

Here, if we consider s in Eq. (6) as a “real” time t , and t in Eq. (6) as a spacial coordinate x , the form of Eq. (6) is similar to the time-dependent Ginzburg–Landau equation for nonconserved systems.⁵⁰ When $W(x, y)$ takes the form of a multiwell function, it is known in general that, after the fast relaxation of the local fluctuation, the global relaxation (the kink motion) takes a very long time.^{51,52} Therefore, in the presence of the hopping motion of $x(t)$ between multiple local minima, the gradient descent method may require a large computational cost.

D. Go-and-Back method

On the basis of perturbation theory and making the proper assumption that $\gamma/\Gamma \ll 1$, we reduce the computational cost as shown below.

To solve the Euler–Lagrange equation

$$\frac{\delta S([x, y]; \Pi)}{\delta x} = \frac{\partial W(x, y)}{\partial x} - \frac{\gamma}{2} \ddot{x} = 0, \quad (8)$$

we decompose the equation into two first-order ordinary differential equations by introducing $v(t)$ which satisfies

$$\gamma \dot{x} = -[V_x^{\text{eff}}(x) + U_x(x, y)] + v. \quad (9)$$

$v(t)$ is a deterministic variable which mimics the random force $\xi(t)$ [see Eq. (1)]. Substituting Eq. (9) into Eq. (8) and using Eq. (2), we finally get

$$\gamma \dot{v} = [V_{xx}^{\text{eff}}(x) + U_{xx}(x, y)]v - k_B T [V_{xxx}^{\text{eff}}(x) + U_{xxx}(x, y)] + \frac{\gamma}{\Gamma} U_{xy}(x, y) [U_y(x, y) - U_y(x^*, y)] + \frac{\gamma}{\Gamma} U_{xy}(x, y) \cdot \eta^*, \quad (10)$$

where \cdot represents Itô-type stochastic calculi⁵³ and x^* represents the true position of x at time t . Since $[y]$ is realized under the true $[x]$, we must use x^* instead of x in Eq. (2).

Here, unknown variables x^* and η^* are involved in the third and the fourth terms in the right-hand side of Eq. (10). Fortunately, the contribution of these two terms are negligible. First, $[U_y(x, y) - U_y(x^*, y)]$ is statistically close to 0 if $[x] \simeq [\hat{x}]$. Similarly, $\langle U_{xy}(x, y) \cdot \eta^* \rangle = 0$. This calculation is independent of $[x]$. Moreover, considering the typical case of single-molecule experiments, we can naturally assume $\gamma/\Gamma \ll 1$.

Therefore, in principle, we may obtain a good approximate solution of $[\hat{x}]$ by solving the following differential equations simultaneously:

$$\gamma \dot{x} = -G_x(x, y) + v, \quad (11)$$

$$\gamma \dot{v} = G_{xx}(x, y)v - k_B T G_{xxx}(x, y), \quad (12)$$

where $G(x, y) \equiv V^{\text{eff}}(x) + U(x, y)$. Here, Eq. (9) is rewritten as Eq. (11), and Eq. (10) is approximated as Eq. (12).

However, a difficulty remains: $G_{xx}(x, y)$ on the right-hand side of Eq. (12) is positive in the typical case when $x(t)$ is fluctuating around the bottom of the energy potential. Thus, as soon as we numerically integrate Eqs. (11) and (12), $v(t)$ will destabilize and finally diverge. In contrast, if we try to solve these equations in the reverse-time direction, $x(t)$ will be instantly destabilized due to the same problem.⁵⁴

To stably solve Eqs. (11) and (12), we need to solve Eq. (11) in the forward-time direction and Eq. (12) in the reverse-time direction. Although it is impossible to solve Eqs. (11) and (12) simultaneously, if the approximate solution of $v(t)$ has been obtained, the approximate solution of $x(t)$ can be calculated by substituting approximated $v(t)$ into Eq. (11) and integrating Eq. (11) in the forward-time direction. By contrast, if the approximate solution of $x(t)$ has been obtained, the approximate solution of $v(t)$ can be calculated by substituting approximated $x(t)$ into Eq. (12) and integrating Eq. (12) in the reverse-time direction. Here, by use of a perturbation expansion technique and the low-order approximations of $x(t)$ and $v(t)$, we develop a method to solve Eq. (11) and Eq. (12) alternatively and increase the order of the approximations systematically.

To perform a perturbation expansion,⁵⁵ we use that the second term in the right-hand side of Eq. (12) is small.

$$\gamma \dot{v} = G_{xx}(x, y)v - \varepsilon k_B T G_{xxx}(x, y). \quad (13)$$

We expand $x(t)$ and $v(t)$ in terms of power series of ε , and we define the i th order approximation as

$$x_{(i)}(t) \equiv \sum_{j=0}^i \varepsilon^j x_{(j)}(t), \quad (14)$$

$$v_{(i)}(t) \equiv \sum_{j=0}^i \varepsilon^j v_{(j)}(t). \quad (15)$$

Note that the present definition of the i th term differs from the usual definition: the i th order approximation includes all terms from orders 0 to i . This definition is crucial to dramatically simplify the algorithm. In addition, we introduce a proper approximation for the zeroth-order term of $v(t)$

$$v_{(0)}(t) = 0. \quad (16)$$

(For the reason, see Appendix.) Then, once we adopt appropriate boundary conditions for $x_{(i)}(0)$ and $v_{(i)}(\tau)$, the higher order approximate solutions of $[\hat{x}]$ can be successively obtained as follows. (For details of the derivation, see Appendix.)

Go-and-Back method:

- Using $v_{(i)}(t)$, we solve

$$\gamma \dot{x}_{(i)} = -G_x(x_{(i)}, y) + v_{(i)} + O(\varepsilon^{i+1}) \quad (17)$$

from $t = 0$ to $t = \tau$ and obtain $x_{(i)}(t)$.

- Using $x_{(i)}(t)$, we solve

$$\gamma \dot{v}_{(i+1)} = G_{xx}(x_{(i)}, y)v_{(i+1)} - \varepsilon k_B T G_{xxx}(x_{(i)}, y) + O(\varepsilon^{i+2}) \quad (18)$$

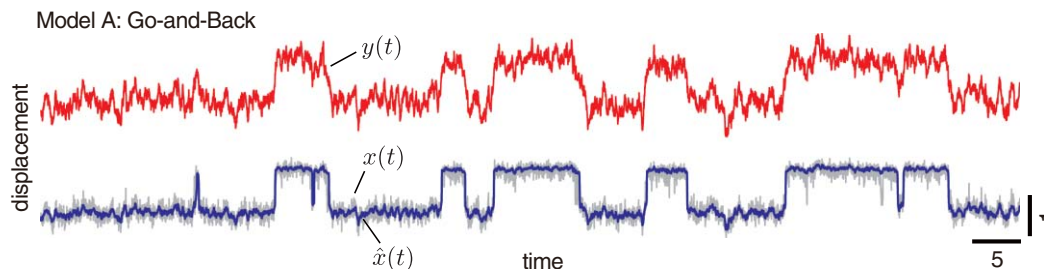


FIG. 2. Estimation result of model A by the Go-and-Back method. (Top) The trajectory of $y(t)$ [red line]. (Bottom) The true trajectory of $x(t)$ [gray line], and the estimated trajectory of $x(t)$, denoted as $\hat{x}(t)$ [blue line]. The iteration number of the optimization process is $i = 50$, and the data length is $\tau = 100$. For the present parameter setting, the relaxation time scale of $x(t)$ is shorter than Δt . Therefore, we use linear-interpolated $[y]$ with the step size $\Delta t/10$ to stably integrate Eqs. (17) and (18).

from $t = \tau$ to $t = 0$ and obtain $v_{(i+1)}(t)$.

- Alternate between Step A and Step B.

III. EXAMPLES

To investigate the practical utility of the Go-and-Back method, we examine the two models illustrated in Fig. 1. We numerically integrate the model equations [Eqs. (1) and (2)] from $t = 0$ to $t = \tau$ and obtain the true trajectory set $[x, y]$. Then, we assume that we can only monitor $[y]$, and estimate $[x]$ from $[y]$.

Figures 2 and 3 display the estimation results of model A and model B by the Go-and-Back method, respectively. Although the estimated trajectory of $x(t)$, denoted as $[\hat{x}]$, does not coincide with the true trajectory $[x]$, $[\hat{x}]$ seems to be a good estimate in both examples. In particular, stepwise motion of $x(t)$ in model B is precisely reproduced from noisy $y(t)$ (Fig. 3).

To apply the Go-and-Back method, a boundary condition is required. (The initial condition is included in the algorithm [Eq. (16)]. The validity will be discussed below.) For both examples (Figs. 2 and 3), $\{x_{(i)}(0), v_{(i)}(\tau)\} = \{y(0), 0\}$ is adopted. To investigate the stability of the Go-and-Back method, we vary the boundary condition of model A at random. As a result, effect of the boundary condition is diminished instantly (Fig. 4). Only less than 0.1% of the total length

of the trajectory at both ends is affected by the boundary condition. Thus, the estimates are almost uniform against the choice of the boundary condition. The Go-and-Back method allows to expect the diminishing time scale from Eq. (11) by the second order approximation on $G(x, y)$ around the bottom of the effective potential. For the present model, the diminishing time scale is approximated as $\tau_{\text{dim}} \sim 0.001$ (Ref. 56), which is almost consistent with the numerical results (Fig. 4).

We also examine the validity of the initial condition adopted in the algorithm. We apply noise to the original initial condition [Eq. (16)] and evaluate the dependency on the estimate. As a result, the original estimate and the estimate obtained from the noise-added initial condition are almost completely overlapped (Fig. 5, see also Fig. S1 in Ref. 57). The difference is 10^{-4} on average, which is 10^4 times smaller than the distance between two stable states, and 10^3 times smaller than the thermal fluctuation of $x(t)$ (Ref. 58). We note that the Go-and-Back method bases on the Euler-Lagrange equation [Eq. (8)], which means we cannot guarantee that the obtained estimate is the global minimum. However, such a robustness against the choice of the initial condition suggests that the estimate may be the global optimum solution at least in the present model. To conclude, the Go-and-Back method incorporates the appropriate initial condition that yields stable solution, and the method is quite robust against the boundary condition.

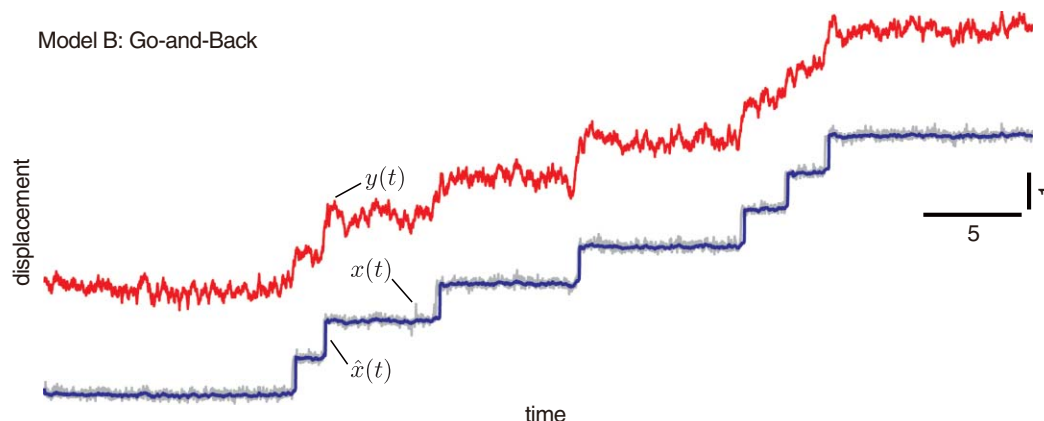


FIG. 3. Estimation result of model B by the Go-and-Back method. (Top) The trajectory of $y(t)$ [red line]. (Bottom) The true trajectory of $x(t)$ [gray line] and the estimated trajectory $\hat{x}(t)$ [blue line]. $i = 50$ and $\tau = 50$. For the present parameter setting, the relaxation time scale of $x(t)$ is shorter than Δt . Therefore, we use linear-interpolated $[y]$ with to stably integrate Eqs. (17) and (18).

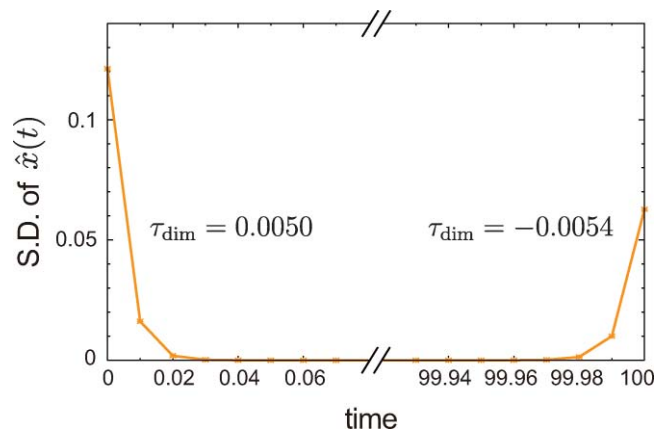


FIG. 4. Effect of the boundary condition. The boundary condition $\{x_{(i)}(0), v_{(i)}(\tau)\}$ of model A is varied randomly ($n = 20$), and the standard deviations of $\hat{x}(t)$ at each t are plotted. The diminishing times at both boundaries are evaluated by fitting exponent functions. $x_{(i)}(0)$ is varied Gaussian noise with mean $y(0)$, SD = 0.5 [half length of the two stable states]. $v_{(i)}(\tau)$ is varied zero-mean Gaussian noise with the variance $2\gamma k_B T$ [the same power of $\xi(t)$].

Next, we compare the Go-and-Back method with the conventional gradient descent method. We use the same $[y]$ as for the Go-and-Back method (Fig. 2, top) and adopt the forward-time centered-space (FTCS) scheme⁵⁹ for the minimization of $S([x, y]; \Pi)$. To apply FTCS scheme, the initial and the boundary conditions are required. For the boundary condition, we adopt Dirichlet condition $\{\hat{x}_0, \hat{x}_\tau\} = \{y_0, y_\tau\}$, and for the initial condition, we adopt raw data of $[y]$, or the moving-averaged trajectories of $[y]$ with different bin sizes in order to investigate effect of the initial condition.

Figure 6 shows the relaxation property of $S([x, y]; \Pi)$ between the two methods. In the case of the Go-and-Back method, after the quick relaxation the value of S fluctuates slightly around the minimum value. In some cases, we can recognize that the method overcomes large barrier of $S([x, y]; \Pi)$ (see Figs. S2 and S3 in the supplementary

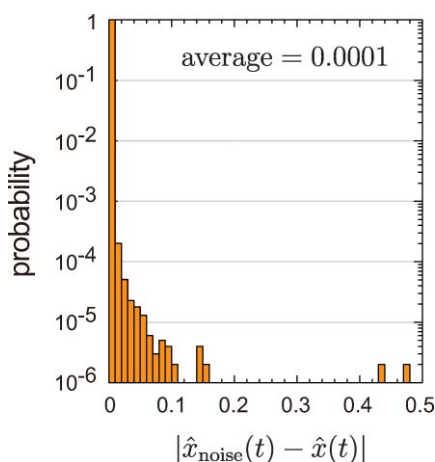


FIG. 5. Effect of the initial condition. Zero-mean white Gaussian noise with the variance $0.25 \times 2\gamma k_B T$ is added to the original initial condition of the Go-and-Back method [Eq. (16)], and the difference between the estimate from the noise-added initial condition $\hat{x}_{\text{noise}}(t)$ and the original estimate $\hat{x}(t)$ is evaluated at each t (10^4 data points). See also Fig. S1 in the supplementary material (Ref. 57).

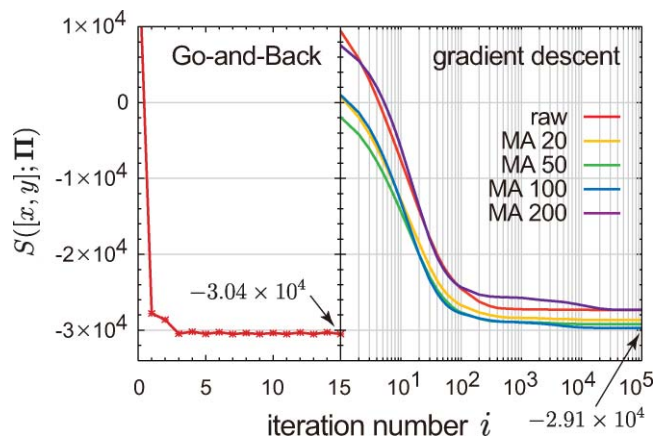


FIG. 6. Relaxation properties of the action functional $S([x, y]; \Pi)$ in the optimization process. (Left) Go-and-Back method. The data is taken from the numerical experiment shown in Fig. 2 (model A). (Right) Gradient descent method. We use the same $[y]$ as for the Go-and-Back method (Fig. 2, top). For the initial condition, we test raw data of $[y]$ and the moving-averaged (MA) trajectories of $[y]$ with bin sizes 20, 50, 100, and 200. (The time lengths are 0.2, 0.5, 1, and 2, respectively.)

material⁵⁷). For the present model, $i = 15$ is enough to converge the solution. Compared to the Go-and-Back method, as one expects, the gradient descent method requires too many iteration steps for the relaxation. (The relaxation time is 10^4 times slower.) Moreover, the converged value of $S([x, y]; \Pi)$ in the case of the gradient descent method is strongly dependent on the initial condition. In the present case, roughly smoothed data (MA 50) yields the smallest value of $S([x, y]; \Pi)$, but it is still a little bit larger than that of the Go-and-Back method. Such a result implies that the optimization processes are trapped at the local minima. Indeed, $\hat{x}(t)$ is varied among the initial conditions. Two examples are shown in Fig. 7. Even the best optimized solution of the gradient descent method (MA 50), the motion of $x(t)$ cannot be precisely estimated (Fig. 7, bottom).

We also apply the gradient descent method to model B and obtain features similar to those observed for model A (see Figs. S3 and S4 in the supplementary material in Ref. 57).⁶⁰

IV. CONCLUSION AND DISCUSSION

We developed a method to estimate the most probable trajectory of the hidden variable from the trajectory of the probe particle. The method naturally incorporates Langevin dynamics. Therefore, although the difficulty of the model settings still remains, if we carefully choose the model, we may extract much more information from experimental data than using the conventional step analysis.

By use of simple models of single-molecule experiments, we numerically verify that the proposed method provides us reasonable estimates. Comparing to the conventional gradient descent method, our proposed method can successfully reduce the computational cost more than 10^3 -fold. In the case of the gradient descent method, the choice of the initial conditions is crucial for obtaining accurate estimates: the wrong choice leaves the optimization process trapped at the local minima. Although we can adopt several techniques, for example, simulated annealing, to overcome this problem,

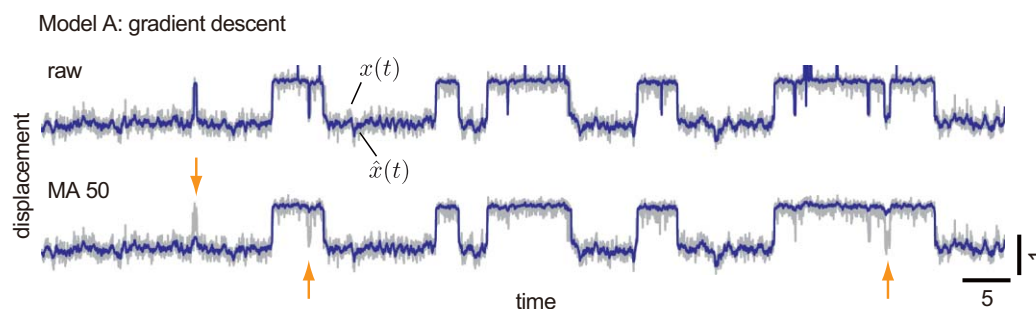


FIG. 7. Two examples of the estimation results of model A by the gradient descent method. The true trajectories of $x(t)$ [gray line], and the estimated trajectories $\hat{x}(t)$ [blue line] are shown. (Top) The raw data of $[y]$ is adopted for the initial condition. (Bottom) A moving-averaged trajectory of $[y]$ with bin size 50 is adopted for the initial condition. The orange arrows show the regions where $x(t)$ cannot be precisely estimated. In both figures, $i = 10^5$.

case-by-case treatment is required to improve efficiency. In contrast, our method naturally incorporates an appropriate initial condition. Besides the simplicity of our coding, the proposed method only requires a boundary condition, and the results are quite robust with respect to this choice. Therefore, our method is easy to use, which is important for general users.

As we mentioned above, the present model has several limitations to apply actual experiments. In the present model, it is assumed that the friction coefficient of protein denoted by γ is independent of $x(t)$. This term originates from the internal friction of protein and the viscous friction between the protein and the medium. In general, the internal friction should be varied along the reaction coordinate. Indeed, position-dependent coefficients have been obtained from model systems and several kinds of proteins by simulation.^{61–64} Therefore, although the constant approximation is not valid for all kinds of proteins, our assumption may be proper for the protein that has a large globular domain and the entire domain tilts or shifts in its structural transition because the internal friction is negligible compared to the large viscous friction.

In addition, the present model assumes that the effective energy potential of the protein $V^{\text{eff}}(x)$ is time independent. However, recent single-molecule studies on cholesterol oxidase⁶⁵ and β -galactosidase⁶⁶ showed direct evidence that the enzymatic turnover events are not a Markovian process but correlated with the previous history. The result indicates that the protein has slow conformation fluctuations and this time dependence should be appeared in $V^{\text{eff}}(x)$. Although, as far as we know, the similar slow dynamics has never been observed in the motor proteins, one has to verify in advance whether such a slow dynamics presents in the protein of interest. By contrast, if the substrate of the enzymatic reaction is not abundant in the medium so that the rate limiting step of the turnover is substrate binding, at least, on and off states should be incorporated into the model. Namely, $V^{\text{eff}}(x)$ should have two (or more) states and switch stochastically.^{46,47,67,68} For this case, new efficient estimation algorithm must be developed on the basis of the switching state model.

In actual applications of our model to experiments, we must estimate the parameters of the entire model in advance. We have proposed a general framework of parameter estimation in the presence of hidden degrees of freedom.³⁶ Accord-

ing to this framework, which is based on Bayesian statistics, we can precisely estimate the model parameters by maximizing the marginalized path probability with respect to the parameters. The marginalized path probability, called a marginal likelihood, is calculated by integrating all possible trajectories of the hidden variables. In this calculation, the DRP, or the MAP estimator in the language of Bayesian statistics, plays the central role. For the analytical approach, the Wentzel–Kramers–Brillouin approximation around the MAP estimator may work well when the temperature of the system is sufficiently small.^{36,39} For the numerical approach, we can utilize the MAP estimator for the initial condition of the Markov-chain Monte Carlo method.^{31,69} By the use of the proposed method, the next step is to investigate how effectively we can obtain reasonable estimates of the model and the hidden trajectories simultaneously.

ACKNOWLEDGMENTS

The authors thank M. Y. Matsuo and S. Toyabe for fruitful discussions. The authors also thank M. Opper, M. Sano, M. Ichikawa, and K. Yoshikawa for helpful comments, and K. Shioguchi, M. Sugawa, T. Nishizaka, and T. Okamoto for discussions on single-molecule experiments. This work was supported from the JSPS Research Fellowships for Young Scientists, Grant No. 20-4954 (to M.M.), and a grant from MEXT, Grant No. 20740239 (to T.H.).

APPENDIX: PERTURBATION EXPANSION

We expand $x(t)$ and $v(t)$ as

$$x(t) = \sum_{i=0}^{\infty} \varepsilon^i x^{(i)}(t), \quad (\text{A1})$$

$$v(t) = \sum_{i=0}^{\infty} \varepsilon^i v^{(i)}(t), \quad (\text{A2})$$

where we denote the i th order terms of $x(t)$ and $v(t)$ as $x^{(i)}(t)$ and $v^{(i)}(t)$, respectively, in order to distinguish from $x_{(i)}(t)$ [Eq. (14)] and $v_{(i)}(t)$ [Eq. (15)].

The zeroth order terms become

$$\gamma \dot{x}^{(0)} = -G_x(x^{(0)}, y) + v^{(0)}, \quad (\text{A3})$$

$$\gamma \dot{v}^{(0)} = G_{xx}(x^{(0)}, y) v^{(0)}. \quad (\text{A4})$$

$G_{xx}(x^{(0)}, y) > 0$ in typical cases, and thus $v^{(0)}(t)$ will instantly destabilize when we solve Eq. (A4) in the forward-time direction. In contrast, if we solve Eq. (A4) in the reverse-time direction, it is expected that $v^{(0)}(t)$ quickly relaxes to around 0. Therefore, we can naturally assume the approximation $v^{(0)}(t) = 0$. Since $v^{(0)}(t)$ is identical to $v_{(0)}(t)$, we introduce the approximation Eq. (16) in the main text.

Owing to this assumption, we can stably solve $x^{(0)}(t)$ in the forward direction by adopting the initial condition on $x^{(0)}(0)$.

The first-order terms become

$$\gamma \dot{x}^{(1)} = -G_{xx}(x^{(0)}, y)x^{(1)} + v^{(1)}, \quad (\text{A5})$$

$$\gamma \dot{v}^{(1)} = G_{xx}(x^{(0)}, y)v^{(1)} - k_B T G_{xxx}(x^{(0)}, y). \quad (\text{A6})$$

It is noteworthy that $x^{(1)}(t)$ is not included in Eq. (A6), and thus we can stably solve Eq. (A6) in the reverse direction. Then, by substituting $v^{(1)}(t)$ into Eq. (A5), we obtain $x^{(1)}(t)$.

Similarly, the second-order terms become

$$\gamma \dot{x}^{(2)} = -\frac{1}{2} G_{xxx}(x^{(0)}, y)[x^{(1)}]^2 - G_{xx}(x^{(0)}, y)x^{(2)} + v^{(2)}, \quad (\text{A7})$$

$$\gamma \dot{v}^{(2)} = G_{xxx}(x^{(0)}, y)x^{(1)}v^{(1)} + G_{xx}(x^{(0)}, y)v^{(2)} - k_B T G_{xxxx}(x^{(0)}, y)x^{(1)}. \quad (\text{A8})$$

Again, $x^{(2)}(t)$ is not included in Eq. (A8). Therefore, we obtain $v^{(2)}(t)$ and $x^{(2)}(t)$ by the use of the lower order solutions. In this way, we can systematically calculate the higher order terms.

However, as the order number increases, the analytical solution becomes complicated. We further develop the method to simplify the algorithm.

We introduce Eqs. (14) and (15). In these definitions, $x_{(0)}(t)$ and $v_{(0)}(t)$ are identical to $x^{(0)}(t)$ and $v^{(0)}(t)$, respectively.

Next, combining Eqs. (A3) and (A5) leads to

$$\gamma \dot{x}_{(1)} = -G_x(x_{(1)}, y) + v_{(1)} + O(\varepsilon^2), \quad (\text{A9})$$

and combining Eqs. (A4) and (A6) leads to

$$\gamma \dot{v}_{(1)} = G_{xx}(x_{(0)}, y)v_{(1)} - \varepsilon k_B T G_{xxx}(x_{(0)}, y). \quad (\text{A10})$$

Similarly,

$$\gamma \dot{x}_{(2)} = -G_x(x_{(2)}, y) + v_{(2)} + O(\varepsilon^3), \quad (\text{A11})$$

$$\gamma \dot{v}_{(2)} = G_{xx}(x_{(1)}, y)v_{(2)} - \varepsilon k_B T G_{xxx}(x_{(1)}, y) + O(\varepsilon^3). \quad (\text{A12})$$

In this manner, we obtain the general forms, Eqs. (17) and (18).

¹K. Svoboda, C. F. Schmidt, B. J. Schnapp, and S. M. Block, *Nature (London)* **365**, 721 (1993).

²H. Noji, R. Yasuda, M. Yoshida, and K. Kinosita, Jr., *Nature (London)* **386**, 299 (1997).

³M. Rief, R. S. Rock, A. D. Mehta, M. S. Mooseker, R. E. Cheney, and J. A. Spudich, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 9482 (2000).

⁴R. Yasuda, H. Noji, M. Yoshida, K. Kinosita, Jr., and H. Itoh, *Nature (London)* **410**, 898 (2001).

⁵W. J. Greenleaf, M. T. Woodside, and S. M. Block, *Annu. Rev. Biophys. Biomol. Struct.* **36**, 171 (2007).

⁶K. Shiroguchi and K. Kinosita, Jr., *Science* **316**, 1208 (2007).

⁷K. Shiroguchi, H. F. Chin, D. E. Hannemann, E. Muneyuki, E. M. De La Cruz, and K. Kinosita, Jr., *PLoS Biol.* **9**, e1001031 (2011).

⁸H. Ueno, S. Nishikawa, R. Iino, K. V. Tabata, S. Sakakihara, T. Yanagida, and H. Noji, *Biophys. J.* **98**, 2014 (2010).

⁹L. Nugent-Glandorf and T. T. Perkins, *Opt. Lett.* **29**, 2611 (2004).

¹⁰D. E. Smith, S. J. Tans, S. B. Smith, S. Grimes, D. L. Anderson, and C. Bustamante, *Nature (London)* **413**, 748 (2001).

¹¹S. Uemura and S. Ishiwata, *Nat. Struct. Biol.* **4**, 308 (2003).

¹²H. Itoh, A. Takahashi, K. Adachi, H. Noji, R. Yasuda, M. Yoshida, and K. Kinosita, Jr., *Nature (London)* **427**, 465 (2004).

¹³T. Watanabe-Nakayama, S. Toyabe, S. Kudo, S. Sugiyama, M. Yoshida, and E. Muneyuki, *Biochem. Biophys. Res. Commun.* **366**, 951 (2008).

¹⁴J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco, Jr., and C. Bustamante, *Science* **296**, 1832 (2002).

¹⁵S. Toyabe, T. Okamoto, T. Watanabe-Nakayama, H. Taketani, S. Kudo, and E. Muneyuki, *Phys. Rev. Lett.* **104**, 198103 (2010).

¹⁶E. Geva and J. L. Skinner, *Chem. Phys. Lett.* **288**, 225 (1998).

¹⁷A. M. Berezhkovskii, A. Szabo, and G. H. Weiss, *J. Phys. Chem. B* **104**, 3776 (2000).

¹⁸J. Cao, *Chem. Phys. Lett.* **327**, 38 (2000).

¹⁹J. B. Witkoskie and J. Cao, *J. Chem. Phys.* **121**, 6361 (2004a).

²⁰O. Flomenbom, J. Klafter, and A. Szabo, *Biophys. J.* **88**, 3780 (2005).

²¹I. V. Gopich and A. Szabo, *J. Chem. Phys.* **124**, 154712 (2006).

²²J. B. Witkoskie and J. Cao, *J. Chem. Phys.* **121**, 6373 (2004b).

²³J. B. Witkoskie and J. Cao, *J. Phys. Chem. B* **112**, 5988 (2008).

²⁴A. Baba and T. Komatsuzaki, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 19297 (2007).

²⁵A. Baba and T. Komatsuzaki, *Phys. Chem. Chem. Phys.* **13**, 1395 (2011).

²⁶S. Chung and R. Kennedy, *J. Neurosci. Methods* **40**, 71 (1991).

²⁷X. Nan, P. A. Sims, P. Chen, and X. S. Xie, *J. Phys. Chem. B* **109**, 24220 (2005).

²⁸J. W. J. Kerssemakers, E. L. Munteanu, L. Laan, T. L. Noetzel, M. E. Janson, and M. Dogterom, *Nature (London)* **442**, 709 (2006).

²⁹B. C. Carter, M. Vershinin, and S. P. Gross, *Biophys. J.* **94**, 306 (2008).

³⁰B. Bozorgui, K. Shundyak, S. Cox, and D. Frenkel, *Eur. Phys. J. E* **31**, 411 (2010).

³¹C. Dellago, P. G. Bolhuis, F. S. Csajka, and D. Chandler, *J. Chem. Phys.* **108**, 1964 (1998).

³²P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler, *Annu. Rev. Phys. Chem.* **53**, 291 (2002).

³³E. Autieri, P. Faccioli, M. Sega, F. Pederiva, and H. Orland, *J. Chem. Phys.* **130**, 064106 (2009).

³⁴P. Faccioli, M. Sega, F. Pederiva, and H. Orland, *Phys. Rev. Lett.* **97**, 108101 (2006).

³⁵M. Sega, P. Faccioli, F. Pederiva, G. Garberoglio, and H. Orland, *Phys. Rev. Lett.* **99**, 118102 (2007).

³⁶M. Miyazaki and T. Harada, *J. Chem. Phys.* **134**, 085108 (2011).

³⁷L. Onsager and S. Machlup, *Phys. Rev.* **91**, 1505 (1953).

³⁸S. Machlup and L. Onsager, *Phys. Rev.* **91**, 1512 (1953).

³⁹K. L. C. Hunt and J. Ross, *J. Chem. Phys.* **75**, 976 (1981).

⁴⁰K. Adachi, K. Oiwa, T. Nishizaka, S. Furuike, H. Noji, H. Itoh, M. Yoshida, and K. Kinosita, Jr., *Cell* **130**, 309 (2007).

⁴¹K. Hayashi, H. Ueno, R. Iino, and H. Noji, *Phys. Rev. Lett.* **104**, 218103 (2010).

⁴²Q. Cui, G. Li, J. Ma, and M. Karplus, *J. Mol. Biol.* **340**, 345 (2004).

⁴³*Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems*, 1st ed., edited by Q. Cui and I. Bahar (Chapman & Hall/CRC, Boca Raton, FL, 2006).

⁴⁴Y. Togashi and A. S. Mikhailov, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 8697 (2007).

⁴⁵Y. Togashi, T. Yanagida, and A. S. Mikhailov, *PLoS Comput. Biol.* **6**, e1000814 (2010).

⁴⁶F. Jülicher, A. Ajdari, and J. Prost, *Rev. Mod. Phys.* **69**, 1269 (1997).

⁴⁷P. Reimann, *Phys. Rep.* **361**, 57 (2002).

⁴⁸H. A. Kramers, *Physica*, **VII**, 284 (1940).

⁴⁹If the system is locally equilibrated, we can adopt the Boltzmann distribution.

⁵⁰J. D. Gunton, M. S. Miguel, and P. S. Sahni, *Phase Transitions and Critical Phenomena*, edited by C. Domb and J. L. Lebowitz (Academic, New York, 1983), Vol. 8.

⁵¹K. Kawasaki and T. Ohta, *Physica* **116A**, 573 (1982).

- ⁵²J. Carr and R. Pego, *Commun. Pure. Appl. Math.* **XLII**, 523 (1989).
- ⁵³R. Mortensen, *J. Stat. Phys.* **1**, 271 (1969).
- ⁵⁴When x is close to the bottom of the effective potential, $-G_x(x, y) \simeq -G_{xx}(x_b, y)(x - x_b)$, where x_b represents the bottom position. Since $G_{xx}(x_b, y)$ is positive, Eq. (11) will be destabilized if the equation is integrated in the reverse-time direction.
- ⁵⁵ ε is a nondimensional parameter, which guarantees the smallness of the second term in the right-hand side of Eq. (12). After the calculation, we substitute $\varepsilon = 1$ into the result.
- ⁵⁶ $\tau_{\text{dim}} \sim \gamma / G_{xx}(x_b) = \gamma / (a_1 + k)$ where x_b represents the bottom position of the effective potential.
- ⁵⁷See supplementary material at <http://dx.doi.org/10.1063/1.3574396> for the supporting figures.
- ⁵⁸The standard deviation of the thermal fluctuation of $x(t)$ is $\sqrt{2k_B T \Delta t / \gamma}$ in the case of discrete t . For model A, SD = 0.2.
- ⁵⁹W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, 3rd ed. (Cambridge University Press, New York, 2007), Chap. 20.
- ⁶⁰The Go-and-Back method is not an algorithm to search the global minimum point of $S([x, y]; \Pi)$ along $[x]$ -space, but systematically calculates the higher order approximation of the solution of the Euler-Lagrange equation [Eq. (8)]. Therefore, a monotonic decrease of $S([x, y]; \Pi)$ does not mean that the estimate gets stuck into the local minimum.
- ⁶¹G. Hummer, *New J. Phys.* **7**, 516 (2005).
- ⁶²R. B. Best and G. Hummer, *Phys. Rev. Lett.* **96**, 228104 (2006).
- ⁶³S. Yang, J. N. Onuchic, and H. Levine, *J. Chem. Phys.* **125**, 054910 (2006).
- ⁶⁴R. B. Best and G. Hummer, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 1088 (2010).
- ⁶⁵H. P. Lu, L. Xun, and X. S. Xie, *Science* **282**, 1877 (1998).
- ⁶⁶B. P. English, W. Min, A. M. van Oijen, K. T. Lee, G. Luo, H. Sun, B. J. Cherayil, S. C. Kou, and X. S. Xie, *Nat. Chem. Biol.* **2**, 87 (2005).
- ⁶⁷T. Harada and S.-i. Sasa, *Phys. Rev. Lett.* **95**, 130602 (2005).
- ⁶⁸T. Harada, *Europhys. Lett.* **70**, 49 (2005).
- ⁶⁹C. M. Bishop, *Pattern Recognition and Machine Learning* (Springer, New York, 2006).